ACT Research and Policy

Research Reports
Educator Reports
Policy Reports
Issue/Information Briefs

Measuring Progress in Core High School Courses: Insights into Value-Added Measures of Teacher Effectiveness

Issue Brief

October 2012

There is great interest in measuring instructional effectiveness for school and teacher accountability, for placing teachers and principals strategically, and for targeting professional development and other supports for educators. Traditionally, teacher quality was largely inferred from the background characteristics of teachers themselves (e.g., teacher experience, advanced degrees, and professional development). In recent years, researchers have developed *value-added* measures that directly measure effectiveness based on assessments of student achievement. These methods use statistical controls for individual prior achievement and background variables to isolate the contributions of teachers and schools to student learning.

Several groups have advocated the use of value-added measures as part of teacher evaluation, including the New Teacher Project, the Bill and Melinda Gates Foundation, Brookings, and Battelle for Kids.[1] In 2009, the federal government offered large education grants to states and required states to use student achievement in teacher evaluations. In 2011, twenty-three states required the use of value-added or growth approaches as part of teacher evaluations.[2]

Value-added critics have argued validly that multiple measures of effectiveness should be used for high-stakes accountability and tenure decisions, because measures of teacher effectiveness based on student assessment do not capture the full impact that teachers have on students.[3] The extent that student performance can be attributed to a teacher is limited by difficulties in isolating teacher effects from school characteristics, students' prior experiences and achievement levels, and other factors beyond the teacher's control.[4] While statistical models can be formulated to control for these factors, they are unlikely to account for all of the relevant variables needed to completely isolate the teacher effect.[5] Measures of principal, school, and district effectiveness have many of the same caveats.

infobrief@act.org for more information or to suggest ideas for future ACT Information Briefs.

ACT®

www.act.org/research

The key issue, however, is not whether value-added measures are perfect, since no evaluation system can measure all aspects of quality without error. Rather, the issue is whether the new measures provide additional valid input to ongoing methods of measuring instructional effectiveness. Teachers are traditionally evaluated by principals during annual (or less frequent) classroom visits. Weisburg et al. found that most principals gave the highest rating to all teachers in the school and 75 percent of teachers received no feedback on how to improve their instruction.[6] Value-added approaches should provide considerably more rigor to these traditional teacher evaluations. Glazerman et al. argue that value-added measures are sufficiently reliable to be used as an element of teacher evaluation.[7] Aside from high-stakes accountability, measures of teacher and school effectiveness can provide valuable information on how much students are progressing toward college and career readiness. Perhaps most importantly, the measures can inform school personnel strategies, such as targeting professional development based on teachers or principals showing less growth and placing teachers or principals with student groups for which they have excelled.

This study demonstrates value-added measures using a large set of data that established baseline growth in eight core high school courses.[8] Actual performance can be compared to expected performance to determine whether students' growth was at, below, or above expectations.

QualityCore and the National Baseline Study

Research has shown that students who take a core high school curriculum (four years of English and three years each of math, science, and social studies) recommended in *A Nation at Risk* tend to be more prepared for college than students who do not.[9] Despite more students taking such courses since that time, only about one-quarter of students who complete a core curriculum are ready for college-level coursework across the four core subject areas.[10] Many courses appear to lack the rigor required to sufficiently prepare students for college.[11]

In response to concerns about the lack of college and career readiness of today's high school graduates, ACT developed QualityCore®, a program designed to help teachers raise the quality of core high school courses. It includes instructional resources, professional development training, formative assessment tools, and end-of-course exams for 12 core courses: English 9, 10, 11, and 12, Algebra I, Geometry, Algebra II, Precalculus, Biology, Chemistry, Physics, and US History.[12]

ACT, with support from the Bill and Melinda Gates Foundation, conducted a national study that established a baseline for the amount of academic growth that occurs in eight of the core courses, including Algebra 1 and Biology.[13] To measure changes in students' knowledge and skills attributable to a given course, the appropriate QualityCore end-of-course examination was administered at the beginning and end of each course. Depending on the grade level of the course, an EPAS[14] assessment (PLAN® or the ACT® tests) was also administered at the beginning and end of each course to measure progress in students' overall level of readiness for college and career in a given subject area. A nationally representative sample of over 35,000 students in 62 schools participated in the study.

While the baseline study used EPAS and QualityCore assessments to measure growth, no other features of QualityCore (e.g., QualityCore instructional resources, professional development training) were implemented. Therefore, it established a growth baseline for core courses, without any instructional interventions. This core course growth baseline was designed to enable evaluation of the effectiveness of instructional improvement programs, including QualityCore.[15]

How This Study Was Conducted

This study used fall EPAS scores, QualityCore end-of-course exam scores, and other relevant variables to calculate expected end-of-course exam scores in two core high school courses, Algebra 1 and Biology. A sample of 3,941 students from 50 high schools was used to calculate expected scores for Algebra 1, and a sample of 6,719 students from 67 high schools was used to calculate expected scores for Biology. This study made use of a regression model to calculate the expected scores.[16] Prior EPAS scores in the same subject area (Mathematics for Algebra 1, Science for Biology) were the primary measures used to determine expected QualityCore end-of-course exam scores. By comparing actual scores to expected scores, it can be determined if students performed as expected, better than expected, or worse than expected in these courses.[17]

Because PLAN and EXPLORE® tests share a common score scale, data from the National Baseline Study—along with additional QualityCore and EXPLORE data collected by ACT after the Baseline Study was conducted—can be used as a basis of comparison for performance observed in the district that is the focus of this study. This study district is in a large metropolitan area that serves over 100,000 students, including high proportions of minority and economically disadvantaged students. High schools in the district administered EXPLORE in the fall of 9th grade, and teachers were trained on how to use QualityCore resources for Algebra 1, English 9, and Biology. Ninth grade students were given the QualityCore end-of-course exams in these courses in the spring.[18]

Value-added models measure the amount of growth that occurred for groups of students, relative to the amount of growth that is expected, based on the pre-course test score, the time between tests, and average pre-course test score in each class (as a measure of peer ability).[19] Value-added scores are commonly calculated as the mean difference between students' actual and expected scores. Value-added scores can be calculated for groups of students under an individual teacher to form a measure of teacher effectiveness. Similarly, scores can be aggregated by schools or districts to form measures of school and district effectiveness. In this study, we aggregated scores for the entire district. Value-added scores can also be aggregated for other student subgroups (e.g., by prior achievement level, race/ethnicity, special needs status) to form subgroup-specific measures of effectiveness. In this study, we also aggregated scores by prior achievement level (EXPLORE score).

What We Found

Tables 1 and 2 below contain summaries of the results by EXPLORE score ranges. The score ranges represent the scores that approximately encompass the bottom 10% (EXPLORE score was <12), 11–50% (12–15), 51–90% (16–20), and top 10% (21–25) of fall-tested 9th graders nationally (ACT, 2007a).[20]

Algebra

In Table 1, the mean expected QualityCore Algebra 1 end-of-course scores are given both overall and for the EXPLORE Mathematics score ranges. Also presented in Table 1 are the number of students at each EXPLORE score range, the actual mean QualityCore Algebra 1 end-of-course score, and the difference between the actual and expected mean (value-added) and associated standard errors. Differences greater than two standard errors above 0.00 indicate that students grew more than expected; differences less than two standard errors below 0.00 indicate that less-than-expected growth occurred. Differences within two standard errors of 0.00 indicate that growth was not significantly different from expected.

Table 1: Measuring Growth in Algebra 1 Courses

EXPLORE Mathematics score	Number of students	QualityCore Algebra 1 end-of-course scores		Value-Added	
		Expected Mean	Actual Mean	Score	Standard Error
<12	727	138.9	141.2	2.3	0.10
12–15	1,353	142.6	143.0	0.4	0.08
16–20	584	145.8	145.1	(0.7)	0.14
21–25	27	149.9	148.1	(1.8)	0.77
Total	**2,691**	**142.4**	**143.0**	**0.7**	**0.06**

Note: negative values are in parentheses.

Of the students who took the QualityCore Algebra 1 end-of-course exam in the district, 2,691 also had matching EXPLORE data from the fall. In this group of students, the mean Algebra 1 end-of-course score that was expected based on the model established from the historical data was 142.4. The actual mean Algebra 1 end-of-course score was 143.0, and the value-added score was 0.7 (note that 143.0-142.4 does not equal the value-added score because of rounding). Because the standard error was 0.06 and the value-added score was greater than two standard errors above 0.00 (0.7/0.06 = 11.7), we conclude that greater-than-expected growth occurred in the math courses implementing QualityCore Algebra 1.

It can be seen in Table 1 that the amount of growth in Algebra 1 differed by achievement level. Students whose EXPLORE Mathematics scores were 15 or below showed greater than expected growth, while students whose EXPLORE Mathematics scores were above 15 showed less than expected growth.

Similar to the analysis for QualityCore Algebra 1, the expected QualityCore Biology end-of-course mean scores are given both overall and for a range of EXPLORE Science scores in Table 2. Of the students who took the QualityCore Biology end-of-course exam in the district, 1,749 also had matching EXPLORE data from the fall. In this group of students, the mean Biology end-of-course score that we expected based on the model established from the historical data was 143.0. Because the actual mean Biology end-of-course score was 144.0, we conclude that greater-than-expected growth occurred in the Biology courses implementing QualityCore Biology.

Like the results for Algebra 1, it can be seen in Table 2 that growth in Biology differed by prior achievement level. Students with the lowest EXPLORE Science scores showed the largest amount of growth. However, unlike the results for Algebra 1, students at all achievement levels, on average, showed greater-than-expected growth in Biology.[21]

Table 2: Measuring Growth in Biology Courses

EXPLORE Science score	Number of students	QualityCore Biology end-of-course scores		Value-Added	
		Expected Mean	Actual Mean	Score	Standard Error
<11	197	137.3	141.4	4.2	0.25
12–15	867	141.9	142.5	0.6	0.12
16–20	662	145.8	145.5	0.7	0.17
21+	23	151.3	154.3	3.1	0.72
Total	**1,749**	**143.0**	**144.0**	**1.0**	**0.10**

Implications

Measuring School Improvement Using Value-Added Models

Value-added analyses can provide a good measuring stick for instructional improvement. Unlike status measures, which are estimates of student achievement at a single point in time, value-added analyses provide evidence of student learning. Because performance at only one point in time is measured, status measures such as proficiency rates do not quantify how much learning occurred.

Historical data, primarily from the QualityCore National Baseline Study, provides a comparison group of fall EPAS test-takers and spring QualityCore test-takers. This design is useful because it allows us to measure the extent to which actual growth differs from expected growth throughout the year in a course or set of courses. Information gleaned from using both assessments in conjunction could be used to make inferences about how much students are learning over time as they progress through coursework and across grade levels.

It is instructive to make growth comparisons against a national baseline standard (as was done here for Algebra 1 and Biology), as well as against a school or district's own baseline (as can be done in subsequent years for any of the QualityCore end-of-course exams). Comparisons against a national baseline can be used to determine how well a district is performing in relation to the rest of the nation. Comparisons across years within a district can be used to determine whether the district is doing a better job of serving the needs of their students over time.

Analyses could be meaningful at the teacher, school, or district levels. At the school or district level, administrators could use the information to assess the extent to which students are making the progress expected of them in various core courses. At the teacher level, the analyses provide information about the effectiveness of individual teachers when certain conditions are present that enhance the meaning and utility of growth measures for measuring effectiveness.

Necessary Conditions for Using Value-Added Models to Attribute Student Growth to Teachers

As noted in the introduction of this paper, it is difficult to directly attribute differences in student growth to individual teachers, schools, or even districts. The primary reason for this is that there are factors beyond the control of school systems and individual teachers that influence student growth. Despite this limitation, value-added models such as those used in this study have promise as an objective way to measure the effectiveness of teachers or school systems. Some of the conditions that enhance the meaning and utility of growth measures for measuring effectiveness include:

1. **Aligned System**. Standards, curriculum, and student assessments should be aligned, providing a cohesive structure within which teachers and school systems can more fairly be assessed. With this condition, there is consistency between what teachers and school systems are measured against and what they are charged with teaching. In the context of measuring effectiveness using QualityCore end-of-course exams, evidence of alignment would include adoption of QualityCore course standards (either in place of, or in addition to, existing course standards), collaborative use of QualityCore resources, and using the QualityCore end-of-course exam as part of students' final course grade. With the adoption of standards that have college and career readiness as an end goal, it is likely that alignment to QualityCore course objectives will become more prevalent.[22]

2. **Aligned Time Frames**. The timing of student assessment is also important; for measuring teacher effectiveness the assessments should span the same time frame as that of the teachers' instruction. Without proper timing of assessments, it is difficult to tease apart the unique contributions of individual teachers affecting student achievement. In the context of measuring effectiveness using QualityCore end-of-course exams for year-long courses, students' prior achievement levels should be measured with assessment from the early fall or prior spring. In this study, we used EPAS assessment from the fall. It would also be possible to use EPAS assessment from the prior spring. Other possibilities include using QualityCore end-of-course assessment from the prior spring (e.g., an English 9 to English 10 growth model) or state assessments from the prior spring or early fall. If available, multiple pre-tests should be used as controls for student learning prior to the start of each course.

3. **Account for a Comprehensive Set of Explanatory Factors**. It is necessary to examine other factors that may affect student performance, such as classroom composition (e.g., classroom/school prior achievement level, classroom/school aggregate socioeconomic status). Accounting for other sources of and influences upon student learning can better isolate the teacher or school system effect. This condition is particularly important when teachers or school systems are compared across many different schooling contexts. In addition, analyses should consider how students select into elective high school classes. For example, if highly-motivated college-bound students tend to select upper-level mathematics courses while less motivated students take lower-level courses, one should take care not to attribute differences in student growth solely to the teachers themselves.
4. **Transparency About the Degree of Uncertainty About the Effectiveness Estimate**. Reporting and understanding standard errors is necessary for proper interpretation of any measure of teacher or school system effectiveness. Like all statistics, teacher and school value-added scores are estimates that are made with some degree of uncertainty. Regardless of whether the estimates are used for high-stakes accountability (e.g., teacher merit pay or principal retaining decisions) or low-stakes reasons (e.g., to assess teachers for targeting professional development), it is important that the user understands the degree of uncertainty about the effectiveness estimate. In general, estimates with greater standard errors (less certainty) should be given less weight for decision-making.
5. **Training and Resources for Interpreting Value-Added Measures and Educator Professional Development**. Districts should provide resources and professional development for principals and teachers to understand what value-added measures mean and how they could be used to improve instruction.

In addition to these conditions, other considerations are important to keep in mind.

- Student assessment should not be the only measure of effectiveness. Other measures should also be taken into account to provide the most information and understanding about the relationships between student growth and teacher effectiveness.[23]
- Co-teaching arrangements provide unique challenges in attributing student growth to a single teacher. In such cases, interpretation at the course or school level is typically most appropriate.
- In some cases, learning standards are distributed across the curriculum and so many teachers among a wide array of subject areas have distributed responsibility for instruction to such standards. In such cases, interpretation at the school level is most appropriate.

Conclusions

There is concern that today's core high school courses are not rigorous enough to sufficiently prepare students for college and careers. The growth analyses presented in this paper can provide some evidence of the extent to which instructional improvements—such as QualityCore—result in improved learning. The QualityCore scores by themselves provide information about the level of knowledge students have about that subject, but when presented in conjunction with information about student growth, they provide a way to measure student learning over time.

There are limitations of the study that should be taken into account when interpreting the results. One limitation involves the extent to which scores from EXPLORE and PLAN can be used interchangeably. Because PLAN—not EXPLORE—was used in the National Baseline Study, the expected end-of-course scores were based on the relationship between fall PLAN scores and spring QualityCore scores. However, EXPLORE scores were used when applying the model to the school district of interest. While the EXPLORE and PLAN assessments cover the same general content areas and share the same score scale, they focus on the knowledge and skills expected of students at different grade levels (i.e., 8th and 9th grade for EXPLORE, 10th grade for PLAN). To the extent to which the two assessments measure different content areas and/or differ in their relative focus on certain content areas, the prediction models may have been different if EXPLORE were used in the National Baseline Study instead of PLAN.[24]

Another limitation concerns differences between the EPAS and QualityCore assessments. For example, EPAS science scores were used to predict performance in the QualityCore end-of-course exam in Biology. The EPAS science assessments measure other science reasoning skills not specific to biology. The same is true with respect to the EPAS mathematics assessments and QualityCore Algebra 1. Further research is needed to investigate the extent to which differences in content area coverage between the assessments used as covariates and those used as the outcome measures affect the resulting value-added estimates.

The results of this study suggest that the district's 9th grade math and biology courses appear to be effective in increasing student learning; in particular, students in biology and low-performing Algebra 1 students had end-of-course exam scores that were greater than what was expected from the baseline study. However, there are other factors that may contribute to student learning, such as other courses, motivation, attendance rates, and family or community influences. This study did take into account students' individual and school-wide prior achievement levels, but there are likely additional factors that were not accounted for in the model. Therefore, while the results of this study provide some evidence supporting the effectiveness of instruction, one should resist fully attributing the student performance to the school system without further consideration of additional evidence.

It is important to note that if a district is performing better than the national average, it doesn't necessarily mean that they are performing at a level high enough to ensure college and career readiness. Nationally, if the overall quality of core high school courses has substantial room for improvement, then schools whose students are showing higher than average growth may still be learning at a rate that is below their potential. Further studies would be useful to assess the extent to which growth in core high school courses in the United States compares to growth in corresponding courses in high-performing countries, and the extent to which growth in core courses varies across our nation's high schools.

This paper illustrates one way in which QualityCore and the EPAS system can provide information about instructional effectiveness within a school or district and whether the instructional improvement programs schools have in place are working. Value-added results should not be the *only* measures of instructional effectiveness. However, such information could be used by schools and districts as a starting point—for prompting further investigation of the factors impacting performance at their schools, for evaluating existing course offerings, and for identifying and targeting professional development opportunities.

Appendix

Table A1 contains descriptive statistics for the sample of students from the district studied. The means are the average EPAS scores, QualityCore end-of-course exam scores, and time span between EPAS and QualityCore testing. The EXPLORE math and science scores are on a scale from 1–25. The QualityCore Algebra 1 and Biology scores are on a scale from 125–175. The national means (SDs) of EXPLORE math and science for fall-tested 9th graders were 16.3 (4.1) and 16.9 (3.2), respectively.[25] The mean EXPLORE math and science scores were 0.7 standard deviations below the national means.

Table A1: Summary of District Samples

QualityCore Exam	Number of Schools	Number of Students	Mean QualityCore Score	Mean EPAS Score	Mean Time Span (Days)
Algebra 1	13	2,691	143.0	13.5	195
Biology	13	1,749	144.0	14.6	197

Table A2 contains descriptive statistics for the sample of students used to establish regression models for the expected QualityCore end-of-course exam scores. The means are the average EPAS scores, QualityCore end-of-course exam scores, and time span between EPAS and QualityCore testing. The PLAN math and science scores are on a scale from 1–32. The mean PLAN math and science scores were < 0.1 and 0.5 standard deviations above the EXPLORE national means for fall-tested 9th graders (recall that EXPLORE and PLAN share a common score scale).

Table A2: Reference Group Samples Used to Establish Models for Expected Scores

QualityCore Exam	Number of Schools	Number of Students	Mean QualityCore Score	Mean EPAS Score	Mean Time Span (Days)
Algebra 1	50	3,941	146.4	16.4	248
Biology	67	6,719	149.9	18.5	243

Table A3 contains the prediction models used to calculate the expected QualityCore end-of-course exam scores. The values in the table are unstandardized regression coefficients and are statistically significant at $\alpha = 0.01$ unless noted.

Table A3: Regression Models to Calculate Expected Scores

QualityCore Exam	Estimated Regression Coefficients			
	Intercept	EPAS	Time Span	Mean EPAS
Algebra 1	125.21	0.8891	0.0207	0.0855*
Biology	116.81	1.1608	0.0037*	0.5768

* Non-significant.

Endnotes

1 Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (2009), accessed July 9, 2012, http://widgeteffect.org/downloads/TheWidgetEffect.pdf; The Bill and Melinda Gates Foundation, *Learning about Teaching. Initial Findings from the Measures of Effective Teaching Project*, accessed February 1, 2011, www.gatesfoundation.org/college-ready-education/Documents/preliminary-findings-research-paper.pdf; Steven Glazerman, Susanna Loeb, Dan Goldhaber, Douglas Staiger, Stephen Raudenbush, and Grover Whitehurst, *Evaluating Teachers: The Important Role of Value-Added* (Washington, DC: The Brookings Institutions, 2010); Kate Kennedy, Mary Peters, and James M. Thomas, *How to Use Value-Added Analysis to Improve Student Learning: A Field Guide for School and District Leaders, Battelle for Kids* (Thousand Oaks: Corwin Press, 2011).

2 National Council on Teacher Quality, *State of the States: Trends and Early Lessons on Teacher Evaluation and Effectiveness Policies*, accessed July 9, 2012, http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf.

3 Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, *Problems with the Use of Student Test Scores to Evaluate Teachers* (Briefing Paper No. 278), (Washington, DC: The Economic Policy Institute, 2010); Eric A. Hanushek and Steven G. Rivkin, *Using Value-Added Measures of Teacher Quality* (CALDER Brief #9), (Washington, DC: Urban Institute, 2010)

4 Stephen W. Raudenbush. “What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice?” *Journal of Educational and Behavioral Statistics* 29 (2004): 121-129. doi: 10.3102/10769986029001121.

5 Linda Darling-Hammond, *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*, accessed May 10, 2011, www.americanprogress.org/issues/2010/10/pdf/teacher_effectiveness.pdf.

6 Weisberg et al., *The Widget Effect*, 2009.

7 Glazerman et al., *Evaluating Teachers*, 2010.

8 ACT. *Raising the Bar: A Baseline for College and Career Readiness in Our Nation's High School Core Courses* (Iowa City, IA: Author, 2012).

9 National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: U.S. Department of Education, 1983).

10 ACT. *ACT Profile Report—National: Graduating Class of 2012* (Iowa City, IA: Author, 2012).

11 ACT. *Rigor at Risk: Reaffirming Quality in the High School Core Curriculum* (Iowa City, IA: Author, 2007).

12 For more information about QualityCore, see http://www.act.org/qualitycore.

13 ACT, *Raising the Bar*, 2012.

14 ACT’s Educational Planning and Assessment System (EPAS) includes three assessments, EXPLORE (for 8th and 9th graders), PLAN (for 10th graders), and the ACT (for 11th and 12th graders). The EPAS system was designed to provide students, teachers, and other stakeholders with information regarding students’ progress toward college and career readiness at three points during their education. The assessments measure achievement of the essential knowledge and skills needed for students to be successful in college and in their careers. The EXPLORE, PLAN, and ACT assessments share a common score scale so that they can be used for longitudinal evaluation of student progress through high school, monitoring whether students are on track and suggesting areas in which they need additional coursework.

15 ACT, *Raising the Bar*, 2012.

16 The model predicted the QualityCore end-of-course exam score for each student, using as predictors the student’s EPAS score in the same subject area, the time span between the fall EPAS test and the spring QualityCore end-of-course exam, and the mean EPAS score for the high school. The same model was used for Algebra 1 and Biology, with the relevant QualityCore and EPAS subject area tests included in the model.

17 Historical data, mostly from the QualityCore National Baseline Study, for students who had taken an EPAS assessment prior to taking the QualityCore end-of-course exam were used to establish expected QualityCore end-of-course exam scores. By using the amount of time between tests as a covariate, the expected scores account for the length of time between the fall test and spring test (more growth is expected as more time passes). (For using time as a covariate, see Allen J., Bassiri, D., and Noble, J. *Statistical Properties of Accountability Measures Based on ACT's Educational Planning and Assessment System (ACT Research Report 2009-1*. Iowa City, IA; ACT, Inc., 2009). Additionally, because of possible peer effects, the expected scores also take into account the mean prior achievement (EPAS Mathematics and Science scores) observed in each high school. (For information on peer effects, see Angrist, J., Lang, K., *How Important Are Classroom Peer Effects? Evidence from Boston's Metco Program, NBER working Paper 9263*, accessed May 10, 2011, http://www.nber.org/papers/w9263. Appendix Table A2 shows the mean baseline end-of-course and pre-class EXPLORE scores of students enrolled in each course.

18 Because QualityCore English 9 was not included in the National Baseline Study, the analysis of instructional effectiveness could only be done for Algebra 1 and Biology. Appendix Table A1 shows mean district end-of course and pre-class EXPLORE scores of students enrolled in these courses.

19 The model here could be expanded to include more controls on student background, including scores on other tests or student demographics. The ultimate goal is to control for the mix of students entering each class, so the growth measure indicate the contribution of the class to student learning by eliminating biases due to the mix of students assigned to a particular class or teacher.

20 ACT. *EXPLORE Technical Manual* (Iowa City, IA: Author, 2007a). The prediction models used to calculate the expected QualityCore end-of-course exam scores are presented in the Appendix, Table A3.

21 As we said earlier, QualityCore provides additional instructional resources relative to the baseline schools. More analysis is needed to isolate whether these resources had a substantive effect on growth or whether the district performed better than the baseline for some other reason.

22 ACT. *The Alignment of Common Core and ACT's College and Career Readiness System* (Iowa City, IA: Author, 2010).

23 For a discussion of research on multiple measures of teacher effectiveness, see the Bill and Melinda Gates Foundation, *Learning about Teaching*, 2010.

24 The EXPLORE Mathematics test contains items assessing four content areas: Pre-Algebra, Elementary Algebra, Geometry, and Statistics/Probability; whereas PLAN Mathematics contains items in Pre-Algebra, Elementary Algebra, Coordinate Geometry, and Plane Geometry. (See ACT, *EXPLORE Technical Manual,* Iowa City, IA: Author, 2007.) Similarly, EXPLORE Science contains items assessing the life sciences, Earth/space sciences, and physical sciences, whereas PLAN Science contains items from biology, the Earth/space sciences, chemistry, and physics (See ACT, Inc. *PLAN Technical Manual*. Iowa City, IA: Author, 2007).

25 ACT, *EXPLORE Technical Manual,* 2007a.

ACT is an independent, not-for-profit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.